

# STANDARDIZED TOOLCHAIN AND MODEL DEVELOPMENT FOR VIDEO QUALITY ASSESSMENT - THE MISSION OF THE JOINT EFFORT GROUP IN VQEG

*Nicolas Staelens<sup>a</sup> Iñigo Sedano<sup>b</sup> Marcus Barkowsky<sup>c</sup> Lucjan Janowski<sup>d</sup>  
Kjell Brunnström<sup>b</sup> Patrick Le Callet<sup>c</sup>*

<sup>a</sup>Ghent University - IBBT, Department of Information Technology, Ghent, Belgium

<sup>b</sup>Acreo AB - NetLab: IPTV, Video and Display Quality, Kista, Sweden

<sup>c</sup>Dept. of Image and Video Communication, IRCCyN, University of Nantes, France

<sup>d</sup>University of Geneva, Switzerland - AGH University of Science and Technology, Poland

## ABSTRACT

Since 1997, the Video Quality Experts Group (VQEG) has been active in the field of subjective and objective video quality assessment. The group has validated competitive quality metrics throughout several projects. Each of these projects requires mandatory actions such as creating a testplan and obtaining databases consisting of degraded video sequences with corresponding subjective quality ratings. Recently, VQEG started a new open initiative, the Joint Effort Group (JEG), for encouraging joint collaboration on all mandatory actions needed to validate video quality metrics. Within the JEG, effort is made to advance the field of both subjective and objective video quality measurement by providing proper software tools and subjective databases to the community. One of the subprojects of the JEG is the joint development of a hybrid H.264/AVC objective quality metric. In this paper, we introduce the JEG and provide an overview of the different ongoing activities within this newly started group.

**Index Terms**— objective video quality metric, subjective video quality assessment, H.264/AVC, standardization, VQEG Joint Effort Group

## 1. INTRODUCTION

Current existing objective video quality metrics are usually proposed and developed by individual organisations e.g. universities or private companies. The whole process of designing such metrics is expensive and time-consuming and usually involves conducting a significant amount of subjective quality assessment experiments in order to obtain valid ground-truth quality ratings. Consequently, the subjective databases that gather these data are usually kept private as they contain valuable information for constructing and validating new objective video quality metrics.

In the VQEG, different organisations e.g. universities, research institutes and private companies, working in the field of both subjective and objective video quality assessment, join

forces in validating competitive objective quality metrics. The results are then given as input into the appropriate study groups of the International Telecommunication Union (ITU), which will formulate Recommendations or standards if judged appropriate.

Following the example of the Joint Video Team (JVT), the joint collaboration between the Video Coding Experts Group (VCEG) and the Motion Picture Experts Group (MPEG) [1] in the standardization of the H.264/AVC video coding standard, some VQEG members also strongly believe in the joint collaboration in both subjective and objective video quality assessment.

In this paper, we provide an overview of the Joint Effort Group (JEG), a new open initiative started within the VQEG in order to facilitate and encourage this joint collaboration of objective video quality metrics. The goal of the JEG is to advance the field of objective and subjective video quality assessment by increasing the scientific knowledge and the general know-how and provide proper software tools and subjective databases to the community. It was decided to focus the initial work to jointly developing a novel no-reference hybrid H.264/AVC objective video quality metric.

The remainder of this paper is structured as follows. We start by providing an overview of the current structure of the VQEG and highlight the new philosophy employed within the JEG. Next, in Section 3, the development cycle for designing, evaluating or validating objective video quality metrics is explained. This section also introduces some abbreviations commonly used within the VQEG and the different test plans. Section 4 provides an overview of the different software tools developed within the JEG and also introduces the XML-based data exchange format used within the group. In Sections 5 and 6, we describe our publicly made available test database and explain our first effort in creating a novel objective video quality metric. Finally, we conclude the paper and present future work.

## 2. STRUCTURE OF VQEG AND JEG

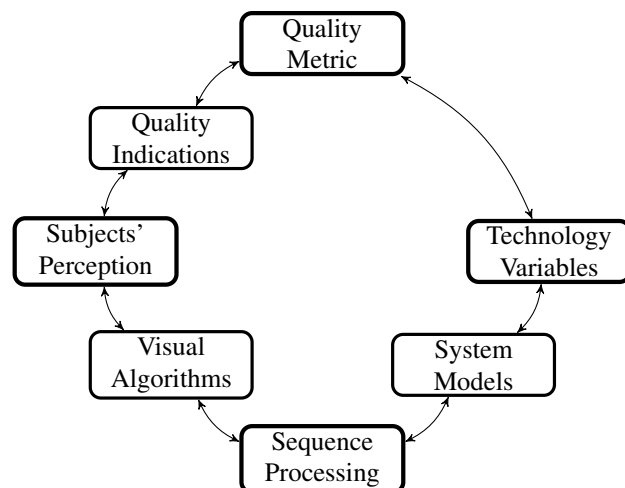
As of 1997, the VQEG has been working towards validating competitive objective video quality metrics for different resolutions up to High Definition (HD) [2, 3, 4] which has resulted in a number of international standards such as ITU-T J.144, ITU-T J.246 and ITU-T J.247. In parallel, VQEG input has led to the adjustment and revisions of some ITU standards on subjective assessment, e.g. ITU-T BT.500 and ITU-T P.910.

The uniqueness of VQEG consists in the wide distribution of domains of the participants which ranges from network providers, helping in the selection of typical transmission conditions, to the manufacturers of objective models. This is complemented by the members of the Independent Laboratory Group (ILG) consisting of several academics, research institutions and even private companies. This mixture allows to identify the current requirements of the industry as well as the most promising solutions.

In order to validate new objective video quality metrics, submitted and developed by different proponents individually, several projects were set up which define the overall scope and validity of the candidate metrics. For example, current ongoing projects include Hybrid Perceptual/Bitstream, Multimedia Phase II and 3DTV. Within each project, a systematic approach is used to validate the submitted objective quality metrics. This includes the formulation of a test plan defining the exact procedure for performing the validation. Each testplan contains the definition of source sequences, typical degradations by coding and transmission in the context of the scope of application, obtaining several subjective test databases in different cultural environments and performing statistical analysis on the results. In order to guarantee a correct and fair validation workflow, the proponents of models are identified and separated from the ILGs.

As a contrast to the other ad-hoc working groups of VQEG, all partners in the Joint Effort Group contribute in a collaborative effort to develop the most suitable, most effective and highest quality video quality assessment strategies. The JEG group is organizing frequent telephone conference calls and face-to-face meetings in the context of the VQEG meetings, which are held at least twice a year. Similar to other groups, e.g. video coding development, the interest of the partners for private intellectual property is not affected.

The first goal of the JEG group is to develop an objective video quality model by using only the information available at the receiver side i.e. no-reference, e.g. inside a set-top box. The bit stream is captured at the network layer and the decoded video signal is stored. The bit stream may be provided to the model in a parsed form which simplifies model development. In order to evaluate and train the objective quality metrics, subjective databases are necessary. A toolchain was created by the JEG in order to ease this training database creation process and up to now, several subjectively evaluated video databases are available for testing. As the JEG is an



**Fig. 1.** Development cycle, composed of different tasks, for fitting the customers' QoE based on [6].

open collaborative approach, the databases are publicly available so that algorithms can be tested and integrated easily into the JEG-Hybrid model.

## 3. VIDEO QUALITY METRIC DEVELOPMENT CYCLE

Like any other software development process, designing, evaluating or validating objective video quality metrics requires a structured methodology composed of a number of consecutive tasks or activities [5]. A typical development cycle for constructing an objective quality metric in the case of image quality was presented by Engeldrum in [6]. Inspired by his proposition, as illustrated in Figure 1, the video quality cycle is presented.

The basic idea of quality metric development presented in Figure 1 is as follows. A sequence is processed by different devices and software libraries such as encoders, streamers and display devices which all have an influence on the obtained quality of the video sequence. Technology variables such as packets inter arrival time, packet loss, I frame size, blocks visibility, and many more are extracted using different system models. Based on those technological variables, Quality of Experience (QoE) can be estimated. Nevertheless, lots of different technological variables can be identified as QoE predictors while only some of them are significant and the relationship may be hard to find for a direct estimation. Therefore, the other path resulting in a QoE model, the left side of Figure 1, is more commonly used. This path starts with investigation the sequences, taking into account the human audio-visual system. This approach helps to identify what a subject perceives. The knowledge of perceived distortions helps to identify the important parameters influencing the video quality. Then, metrics measuring those distortions can be pro-

posed.

As shown in Figure 1, all processes used to build a quality metric are connected by bidirectional links as the obtained results always influence the next steps and indicate possible changes in the previous steps. Moreover, each block represents an expert knowledge which is a large topic by itself. Obviously, significant progress in the domain can be obtained only if the understanding of any part of the graph presented in Figure 1 is shared and can be used as an inspiration for change in the other part of the QoE metric development. Nevertheless, proprietary solutions often use metrics with secret algorithms and even if the algorithms are published, the reasoning for the algorithm development is missing and further improvements are hindered. Therefore, JEG is going to encourage the video community to work more jointly getting solutions closer to the optimum presented in Figure 1.

The first step which makes the collaboration within the video quality community easier is to standardize the input information. Therefore, JEG focuses on building a common platform to inject impairments into the sequences and provide all information on the processed sequences in an easy to read XML format. The toolchain enabling this functionality is presented in the next section.

#### 4. TOOLS IN THE PROCESSING CHAIN OF JEG

Within the JEG, different software tools are developed and used which facilitate the creation of Hypothetical Reference Circuits (HRCs) and corresponding Processed Video Sequences (PVSs). All these software tools are bundled in a Linux-based virtual machine, publicly available on the website of the Tools and Subjective Labs Setup group [7] of VQEG. Furthermore, the JEG also proposed the use of a new XML-based data exchange file format, which captures and combines all available information from the encoded video bit stream, the decoded video sequence and the network level (in case of video streaming). It is important to mention that the tools listed in the following subsections are work in progress. The tools are continuously enhanced and adjusted based on the needs of the JEG.

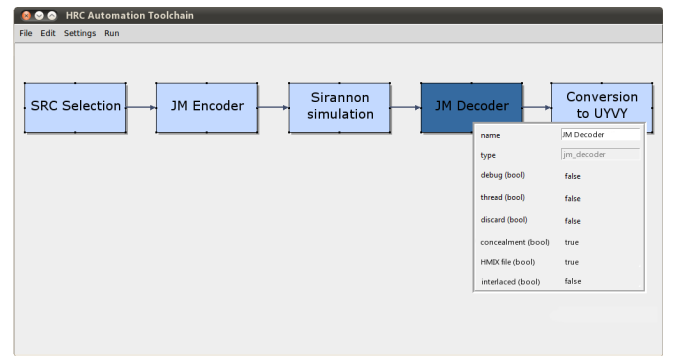
##### 4.1. HRC Automation Toolchain

As explained in the previous section, HRC creation involves selecting, encoding and optionally impairing a series of video sequences in order to produce a set of PVSs. Encoding a video sequence requires specifying the different encoding parameters such as bit rate, quantization parameter, GOP structure and length, . . . Next, impairing the encoded sequence usually includes simulating or emulating video streaming, which also comprises injecting realistic network impairments such as packet loss, and capturing the impaired/degraded video stream. Finally, the PVSs are generated by decoding the impaired video sequences. This whole process is repeated for

each original source sequence and is usually automated using any available programming language.

However, setting up this entire process is time-consuming and programming intensive in order to ensure the correct settings are used for encoding and streaming the video sequences and the PVSs are generated correctly.

In order to facilitate HRC and PVS generation, the JEG is working on an automation toolchain which enables setting up the entire process of PVS generation through a Graphical User Interface (GUI), illustrated in Figure 2, by combining different processing blocks as a graph. Each processing block corresponds with a specific action such as encoding, streaming, impairing, decoding or analysing a source video sequence.



**Fig. 2.** Graphical User Interface for setting up the PVS generation toolchain by combining processing blocks.

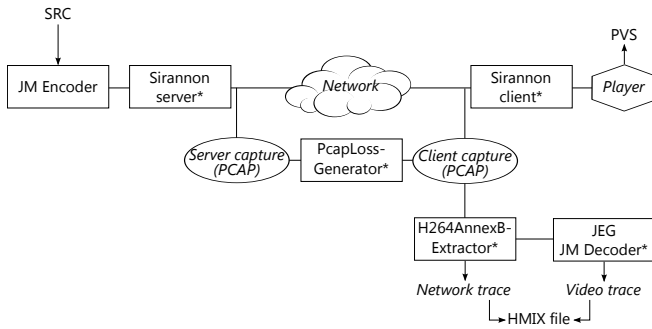
The PVS generation process can also be started and controlled from within this GUI. Internally, the configuration of the toolchain is saved as an XML file and converted to a number of scripts responsible for the execution of a particular processing block. The software can also easily be extended in case processing blocks must be added or adapted (in case, for example, new encoders must be used). As mentioned before, this toolchain makes part of a virtual machine combining all the software tools developed or used by the JEG. In case multiple virtual machines are deployed in the system, the whole process of generating PVSs can be parallelized by distributing the work to all available machines.

The advantage of providing a GUI for HRC and PVS generation is that all internal programming details necessary for deploying and running the scripts are hidden for the user. Using this toolchain also enables researchers to generate PVSs more rapidly without having to spend too much effort in programming their own software.

##### 4.2. Monitoring and gathering information at video and network level

During the PVS generation process, as much information as possible should be collected from the encoded video bit stream,

the decoded video signal and the network level which then serves as input for constructing the hybrid No-Reference (NR) H.264/AVC objective video quality metric. Different software tools are provided with the virtual machine which are used to automatically gather all the required information. Figure 3 illustrates in more details which information is collected, using which available software tools, while streaming and impairing different source sequences for generating corresponding PVSs. The toolchain used for this setup is more detailed than the one depicted in Figure 2 as we now include real video streaming. This setup was also used for generating the example database which will be explained in Section 5. Software tools marked with a star are contributed by JEG members and will also be explained in more details.



**Fig. 3.** Detailed toolchain for generation PVSs and gathering as much information as possible from the video and the network level, using different software tools contributed by JEG members.

At the network level, information is gathered using *Sirannon*<sup>1</sup> [8], *tcpdump* and *H264AnnexBExtractor*. *Sirannon* is a modular multimedia streamer supporting a wide variety of video formats and streaming protocols and can be used for both real-time video streaming and offline simulation. While streaming, *Sirannon* can output a number of trace files containing information about the packetization process and timing information for every packet sent/received. The *tcpdump* program is used for capturing the network packets at server and/or client side and creating a binary capture dump file. Network conditions and impairments can be simulated or emulated by either streaming the video over a real network, by using *Sirannon* as an impairment proxy or by injecting packet losses directly on the server capture file using the *PcapLoss-Generator* program. Next, the client capture file is used as input to *H264AnnexBExtractor* which parses the file (using *TracesPlay* [9]) and reconstructs the (impaired) Annex B compliant video bit stream and outputs information about the streaming process in an XML file. This approach ensures that as much information as possible is captured from the network.

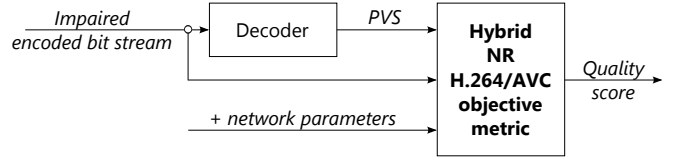
Video level information is extracted while decoding the received (impaired) video bit stream using the *JEG JM De-*

*coder*. The latter is an enhanced version of the JM Reference Software version 16.1 and supports error concealment and the generation of a trace file containing the whole structure of the encoded bit stream up to and including all macroblock coefficients.

Next, we will explain how all this information is efficiently combined into one XML-based trace file and how this facilitates the construction of an objective video quality metric.

### 4.3. XML-based Data Exchange File Format

The JEG is currently focusing on the construction of a hybrid NR H.264/AVC objective video quality metric. This implies that the objective metric takes as input the decoded impaired video signal i.e. the PVS, the corresponding encoded bit stream and information from the network as illustrated in Figure 4.



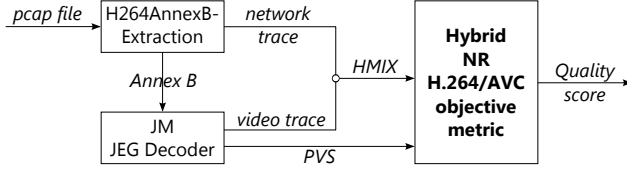
**Fig. 4.** A hybrid NR objective video quality metrics takes as input a PVS, the encoded bit stream and side information from the network.

This approach, however, requires that the objective metric parses the video bit stream and the network parameters itself in order to extract all necessary parameters. This is not a straightforward task. Therefore, the JEG developed a new XML-based data exchange file format, referred to as the Hybrid Model Input XML (HMIX) file, which contains all available information from the encoded (impaired) video bit stream and the network level. Using the XML markup language enables easy processing of the file contents without the need of writing a complete bit stream parser and leads to a faster development of new objective metric building blocks.

As pointed out in the previous subsection, information from the video and the network level is already given as output during PVS generation by the *JEG JM decoder* and the *H264AnnexBExtractor* software tools. Extracting and combining this information into one HMIX file only requires that the captured pcap file from Figure 3 is made available. The HMIX file is then used as input to the hybrid model, as shown in Figure 5, and replaces the impaired encoded bit stream and the network parameters.

From the network level, the HMIX file contains information about each RTP packet and includes, amongst other, the sequence number, timestamp and NAL unit type. This way, information on the streaming process can be obtained. At the video level, detailed information up to the level of the macroblocks, motion vectors and coefficients is outputted to the

<sup>1</sup> Sirannon is formerly know as xStreamer.



**Fig. 5.** Instead of using the PVS and the encoded video sequence, the HMIX file can be used as input to the hybrid NR objective video quality metric which facilitates parsing and processing.

HMIX file. As such, the HMIX file contains all aggregated information which can be parsed and analysed more easily.

## 5. EXAMPLE DATABASE

Currently, an example database consisting of publicly available HMIX files has been created. The sequences which were used to create these HMIX files were all taken from the VQEG HDTV project [4]. So far, the HMIX files corresponding to the HDTVPool2 [10] and HDTVPool3 sequences with encoding-only artefacts have been generated and can freely be downloaded from <http://www.irccyn.ec-nantes.fr/spip.php?article491> and <ftp://intecftp.intec.ugent.be/vqeg/jeg/hmix/hdtvpool3>, respectively. The HMIX files contain large amounts of information in XML format (more than 1 GB for a 10 seconds sequence), varying depending on the number of parameters extracted. Table 1 provides a short description of the different HRCs corresponding to the HMIX files.

**Table 1.** HRCs corresponding to the sequences for which the HMIX files are publicly available.

	HRC	Encoding settings
HDTVPool2	01	QP26
	02	QP32
	03	QP38
	04	QP44
	09	QP26, Rescaled Transmission in 720p
	10	QP38, Rescaled Transmission in 720p
	11	MPEG-2 at QP15, followed by H.264 at QP32
HDTVPool3	04	Bitrate: 15Mbps
	07	Bitrate: 10Mbps

The reader is referred to the VQEG HDTV Final Report [4] for obtaining the individual subjective quality ratings corresponding to the video sequences listed in Table 1.

Interested parties are encouraged to download and experiment with the available files and to provide their feedback on

the obtained results and findings back to the JEG.

## 6. HYBRID VIDEO QUALITY METRIC

The objective of a hybrid model is to estimate the quality as perceived by the end-user. The model can combine information from the HMIX file and the decoded video. For example, a model could locate the pixels affected by a transmission distortion using the information of the HMIX file and evaluate just those pixels in the decoded video. The expected output of such a hybrid model is an estimated Mean Opinion Score (MOS) value. Many interesting parameters related to compression and transmission distortions have already been pointed out in the literature. However, those parameters have to be used and combined in an appropriate way to produce an accurate estimation, which is a very challenging task. Therefore, the JEG is continuously looking for any kind of contributions in order to improve the hybrid model and to encourage joint metric development.

A first objective video quality metric has been developed and implemented which estimates perceived quality as a function of the encoded QP value. Results in [10] show that there is a high linear correlation between the QP value and the subjective quality ratings for H.264/AVC encoded video sequences. The proposed algorithm is implemented using the Python programming language and takes an HMIX file, as described in Section 4.3, as input. In fact, the hybrid model can be programmed in any available language, as long as a standardized interface such as XML is supported. The source code of the algorithm, listed below, shows that the HMIX file can be efficiently processed and analysed using only a limited number of instructions in order to estimate the MOS.

**Listing 1.** Algorithm for estimating MOS based on encoded QP value.

```

#Usage: python hybrid.v1_qp.py [XML file to parse]
import xml.etree.cElementTree as ET
import time
import sys

if len(sys.argv) < 2:
    print "No file name specified as argument"
    sys.exit()
else:
    fileToParse = sys.argv[1]

context = ET.iterparse(fileToParse, events=("start","end"))
context = iter(context)

#get the root element
event, root = context.next()

sumPicturesQP = 0
numPictures = 0

for event, elem in context:
    if event == "end" and elem.tag == "Picture":

        sumMacrosQP = 0
        numPictMacros = 0
        qpList = elem.findall("./Slice/MacroBlock/QP_Y")
        for macroQP in qpList:
            sumMacrosQP = sumMacrosQP+int(macroQP.text)
  
```

```

        numPictMacros = numPictMacros + 1
        meanPictQP = float(sumMacrosQP) / float(numPictMacros)
        sumPicturesQP = sumPicturesQP + meanPictQP
        numPictures = numPictures + 1

    root.clear()

meanSequenceQP = float(sumPicturesQP) / float(numPictures)

estimatedMOS = -0.172 * meanSequenceQP + 9.249
print "Estimated MOS: " + str(estimatedMOS)

```

This model, which has a correlation of 0.956 [10], is only valid for sequences coded with a constant QP value (the formula applies approximately between a QP of 28 and 44) and does not take into account the distortions due to transmission. The algorithm iterates over all the pictures in the video sequence and calculates the average QP value based on the QP\_Y values of each macroblock in that picture. Then, the average QP for the entire sequence is calculated and the MOS is estimated using the following equation:

$$estimatedMOS = -0.172 * meanSequenceQP + 9.249$$

More information on this objective video quality metric can also be found on the JEG wiki page at <http://wiki.vqeg-jeg.org>.

Our preliminary results showed that even for sequences coded with constant bit-rate, the QP seems to provide a better indicator than the actual bit rate value. The next version of the hybrid model could thus focus on the case of variable QP encoding and search for a better indicator instead of using the average QP value.

Using our XML-based HMIX file as input shows that all the required information can easily be extracted without the need for writing an entire bit stream parser. This facilitates the development of new objective video quality metrics.

## 7. CONCLUSION AND FUTURE WORK

With the start of the JEG, the VQEG wants to encourage and facilitate the joint development of objective video quality metrics. A lot of effort has been spent to provide the research and scientific community with proper software tools and subjective test databases. The way of structuring, systematizing and breaking down the developing process will facilitate the understanding and common knowledge about video quality assessment.

The work presented in this paper is a continuous ongoing effort and is based on contributions from different JEG members. As the JEG is an open initiative, any interested party is invited to join the group to share and contribute ideas, expertise, tools, ... and to work in a collaborative approach.

To get in touch with the JEG or to stay informed about ongoing activities and results, subscribe to the JEG mailinglist at <http://www.its.bldrdoc.gov/mailman/listinfo/jeg> or consult the online wiki page, available at <http://wiki.vqeg-jeg.org>, and the discussion forum at <http://forum.vqeg-jeg.org>.

## 8. ACKNOWLEDGMENT

This work was partly conducted for the ICON OMUS project, funded by Interdisciplinary Institute for Broadband Technology (IBBT); the SVC4QoE project funded by the French Direction Générale de la Compétitivité, de l'Industrie et des Services (DGCIS); the INSIGMA project no. POIG.01.01.02-00-062/09 funded by the European Regional Development Fund under the Innovative Economy Operational Programme. Iñigo Sedano would like to thank the ECOC Foundation for financial support through his PhD grant.

## 9. REFERENCES

- [1] Iain E.G. Richardson, *H.264 and MPEG-4 Video Compression: Video Coding for Next Generation Multimedia*, John Wiley & Sons Ltd, 2003.
- [2] VQEG, "Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II," August 2003.
- [3] VQEG, "Final Report from the Video Quality Experts Group on the Validation of Objective Models of Multimedia Quality Assessment, Phase I," March 2008.
- [4] VQEG, "Report on the Validation of Video Quality Models for High Definition Video Content," June 2010.
- [5] M.C.Q. Farias and S.K. Mitra, "A Methodology for Designing No-Reference Video Quality," in *Fourth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM-09)*, January 2009.
- [6] Peter G. Engeldrum, "A theory of image quality: the image quality circle," *Journal of Image Science and Technology*, vol. 48, no. 5, pp. 447–457, September/October 2004.
- [7] "VQEG Tools and Subjective Labs Setup website," <http://vqegstl.ugent.be>, [online].
- [8] A. Rombaut, N. Staelens, N. Vercammen, B. Vermeulen, and P. Demeester, "xStreamer: Modular Multimedia Streaming," in *Proceedings of the seventeenth ACM international conference on Multimedia*, 2009, pp. 929–930.
- [9] "Tracesplay," <http://tracesplay.sourceforge.net/>, [online].
- [10] M. Barkowsky, M.H. Pinson, R. Pépion, and P. Le Callet, "Analysis of Freely Available Subjective Dataset for HDTV Including Coding and Transmission Distortions," in *Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM-10)*, January 2010.